# FRBR Applied to Scientific Data

Joseph A. Hourclé, Wyle Information Systems, <joseph.a.hourcle@nasa.gov>

*The distinction between a creative work and the physical item that contains that work is clearly delineated in FRBR and other research by the Library Science community. A similar confusion exists in the scientific realm between the underlying scientific data and the digital objects that contain those data. We present a similarly scoped reference framework for sensor-based scientific data, drawing on the concepts in FRBR, and compare it with the application of FRBR for cataloging other non-book records.*

Virtual Solar Observatory

Information Systems

## Introduction

The science community has recognized a need for sharing of scientific data, both within a given scientific discipline and across typical scientific boundaries (NASA, 2007). With this sharing, we gain the opportunity for re-use and increased value from a given scientific experiment.

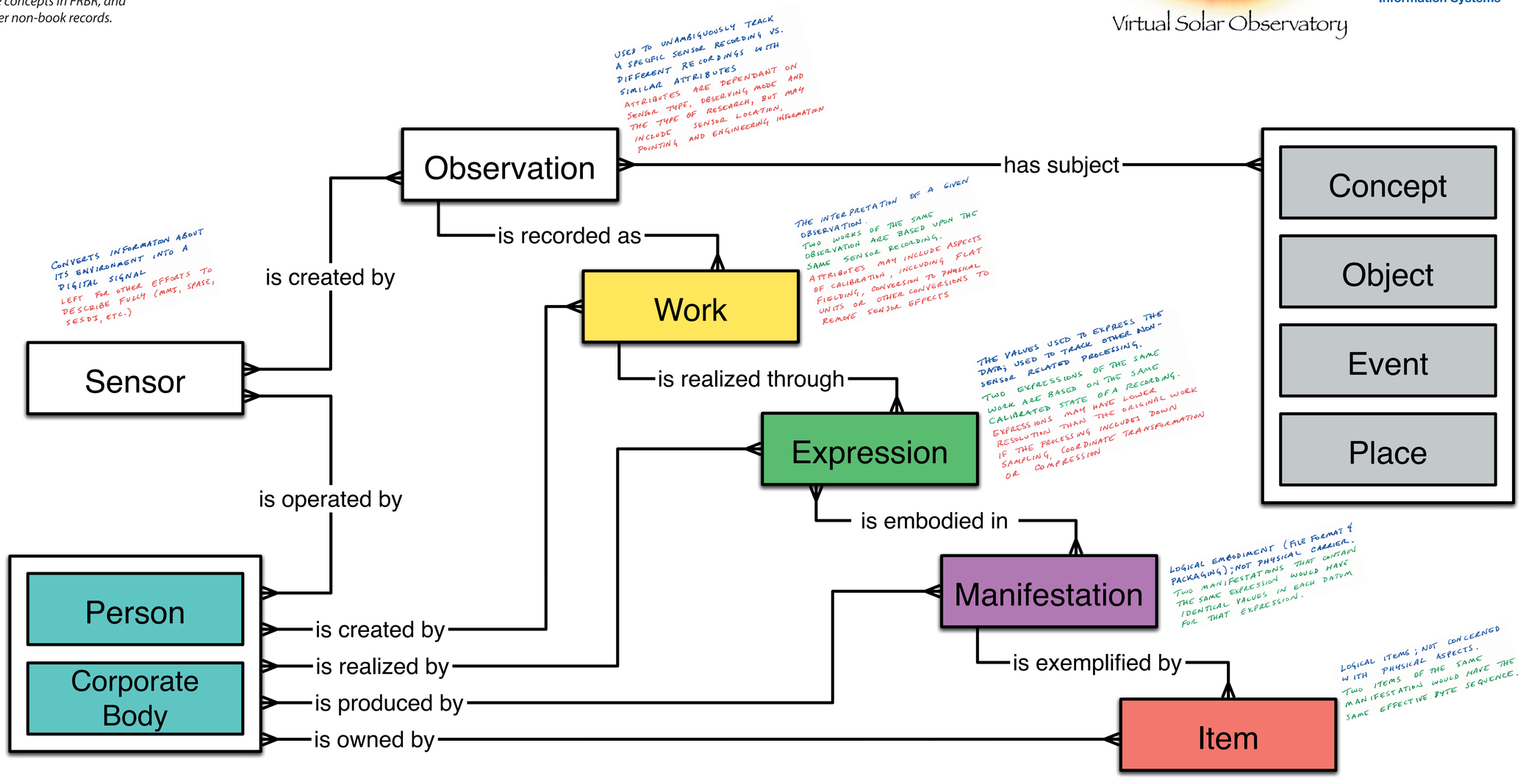
Scientific data may be processed into multiple derivative works, each having a different intended use or audience. The hierarchical relationship between the initial sensor data and the multiple ways in which the data may be translated, packaged and stored is similar to the FRBR (IFLA, 1998) concepts of Work, Expression, Manifestation and Item (Hourclé, 2007). We assume that the catalog's primary audience is the scientists looking for data; we stress the "Functional" aspect of FRBR and the loose interpretations of FRBR may directly conflict with typical FRBR usage. We focus on the FRBR tasks of 'identify' and 'select', which are lacking in the Open Archive Information System [OAIS] Reference Model (CCDS, 2002). The necessary identification tasks include:

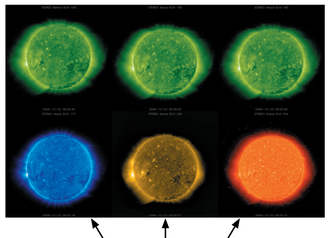
- Which sensor did the data come from?
- Does this data describe an event or object I am interested in?
- Is the data calibrated for the use I want?
- Has the data been reduced in any way?
- Which data package is best for my tools?
- Is a local copy available?
- Can I obtain the data from an alternate location?

An archive may be able to handle some of these questions, but the lack of coordination of identifiers between archives makes it difficult for designers of federated search systems to make decisions about how to present merged search results when there are two files available that seem similar:

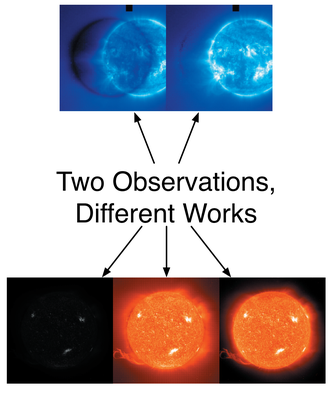
- Should I show both of these records or a prototypical example?
- Can I cluster the records to keep from overwhelming the user?
- At what point is it necessary to disambiguate between two similar items?
- Which object is better to serve the user's needs?

The primary goal of this model is to allow for better differentiation and specificity in searching by decreasing the amount of duplication that is presented to the user. The secondary objective is to build catalogs that more accurately describe the relationships between instances of data, the articles they are used in, and the other supporting objects.

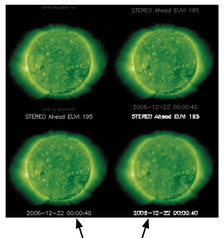


Different Observations

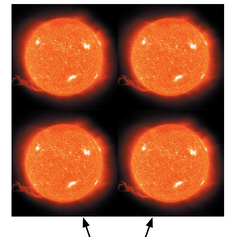


Two Observations, Different Works



Same Work, Different Expressions

Same Expression, Different Manifestations



Same Manifestation, Different Items

## References

See the original paper for the full list of references and URLs.

Consultative Committee for Space Data Systems (2002). *Reference Model for an Open Archival Information System (OAIS).*

Hourclé, J.A. (2007). FRBR in a Scientific Data Context, *Science Archives in the 21st Century*, Adelphi, MD – April 25-26, 2007.

IFLA Study Group on the Functional Requirements for Bibliographic Records, & International Federation of Library Associations and Institutions. (1998). *Functional Requirements for Bibliographic Records: final report*. München: K.G. Saur.

National Aeronautics and Space Administration (2007). NASA *Heliophysics Science Data Management Policy*.

Thank you to Ingbert Floyd, Karen Wickett, Sunny Yoon, Mary Ann Munn and the programmers and scientists on the VSO and SPASE projects for their help with this paper.